DOCUMENT RESUME

ED 289 554 JC 880 011

AUTHOR Boss, Roberta S.
TITLE Formative Evaluation of College Composition: A Formula for Revision and Grading.
PUB DATE [88]
NOTE 31p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Community Colleges; Comparative Analysis; Conventional Instruction; Evaluation Criteria; *Evaluation Methods; Feedback; *Freshman Composition; Grading; *Holistic Evaluation; *Peer Evaluation; Student Attitudes; Teacher Response; Two Year Colleges; *Writing Evaluation; *Writing Improvement

ABSTRACT

A study was conducted at the University of Maryland (College Park, Maryland) in the Fall 1985 15-week semester to test the effects of advanced knowledge of grading criteria on students' writing skills and attitudes in a Freshman Composition class. A holistic grading scale, which was distributed to students as a checklist for revising writing assignments, was developed and coordinated with the assignment sheet for each of six major papers. Students were assigned to either direct-instruction or peer-critique groups. Students in the direct-instruction group submitted their drafts to teachers and received written and in-class feedback. Students in the peer-critique groups had the opportunity to grade sample papers and classmates' drafts using the grading scale. The groups were compared on the basis of grade improvement on pre- and post-tests; the amount of out-of-class time spent by the teachers in responding to individual student papers; and student attitudes toward the grading procedures. The study found no significant differences in grade improvement or student attitude toward grading between the groups, although teachers spent significantly more out-of-class time responding to the papers of the direct-instruction group than to the peer-critique group. Appendixes include a sample assignment sheet and the grading scale/revision checklist. (EJV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED289554

FORMATIVE EVALUATION OF COLLEGE COMPOSITION: A FORMULA FOR REVISION AND GRADING

by

Roberta S. Boss, Ph.D.

Adjunct Professor
Dept. of English and Humanities
Montgomery College
Takoma Park Campus
7600 Takoma Avenue
Takoma Park, Maryland
20912

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roberta S. Boss

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

JC 880 011

BEST COPY AVAILABLE

ABSTRACT

This study tested the effects on student writing of advance knowledge of grading criteria. An instrument was developed combining holisting rating with analytic reading of essays. The six scales, matching the requirements of the semester's major assignments, were used twice: once by students for self- or peer-revision of first drafts and then by teachers for grading final papers. This experiment focused upon the manner of communication of writing goals to students: through direct teacher instruction or through peer critique. There was no significant difference in grade improvement or student attitudes toward grading between the groups, although teacher response time was less than half for the peer group. Five of the six classes showed improvement far beyond that expected through maturation (t-scores well above critical 1.67, $p < .05$). Pre- and posttests were graded holistically by independent raters using ACE guidelines (1985).

SCOPE OF THE PROBLEM

Purposes of Evaluation

Learning theory describes three purposes of evaluation in education: predictive/diagnostic, to screen students for placement in a program (as college entrance examinations), or to identify problems at the start of instruction; formative, which provides immediate and usable feedback to both teachers and students throughout the semester; and summative, to measure competency at the completion of a unit or course of study (Bloom, Hastings & Madaus, 1971). Pre-test/posttest comparisons are both predictive and summative, in that they examine entry versus exit skills.

For administrators of school systems or researchers, a combination of predictive and summative measures is sufficient to compare large groups in achievement over time. Diagnostic tests have their place in the classroom, to aid in prescriptive teaching and remediation. However, Charney (1984) claims that amassing summative statistics (recording grades assigned on a curve) can serve only to rank students in a class, not help a teacher identify and solve an individual student's writing problems.

Many studies which appear to address formative evaluation have been concerned with the manner of grade reporting. For example, Stevens (1973) studied the effects of positive and negative comments on student papers; Bata (1972) found that specific criticism and corrections "helped more" than general overall suggestions; Stanton (1974) found no differences in grade improvement from "checklist, instruction, and questions/feedback," except that the checklist seemed to help teachers be "more reliable [in grading]."

Methods of Assessment

To assist administrators in obtaining predictive and summative data on large numbers of student compositions, holistic rating was developed. It is a global, impressionistic appraisal of the quality of student essays. Cooper and Odell (1977) describe the method as anything short of counting linguistic features of text. Each paper is compared to others in the set or to "benchmark" models which have been pre-graded (norm-referencing). The American Council on Education, which is investigating the feasibility of adding an essay to the GED English examination, has published guidelines for new raters (1985). Six levels are described (not five, to avoid the "average" score) and papers are divided into upper and lower level, then subdivided into "high, middle, and low" to score. Papers receive only the numerical score and remain unmarked and uncorrected. Claims of high inter-rater correlations (reliability) are attributed to "peer pressure, [leader]-monitoring, and rating speed" (Charney, 1984). Multiple readings and smaller ranges of scores also serve to standardize the scoring.

The California Essay Scales (1960) provide readers with six models chosen from 561 expository essays written by 12th-graders. The papers are ranked, with correction symbols, marginal notes, and critical comments on each. Such feature analysis is not characteristic of holistic scoring. Other instruments to evaluate writing are contained in a handbook (Fagan, Cooper & Jensen, 1975) for researchers. These describe broad criteria, often including norming data or complex scoring and weighting directions.

Modified holistic scoring. When analytic reading of student essays on the SAT's proved too time-consuming, the Educational Testing Service sought a means of determining writing proficiency more quickly and reliably (Godshalk, Swinford & Coffman, 1966). Although objective, multiple-choice items could be machine-scored, their validity was questioned, in that they tested only "fragmentary" prerequisite skills, and not a student's writing ability (Lloyd-Jones, 1981). For example, the verbal analogies section measures verbal fluency and possibly reasoning, but it is not a "direct measure" of written expression.

The Composition Evaluation Scales (Diederich, 1961) were developed after a study in which 53 "expert readers" were asked to rate 300 college essays on a scale of 1-9.

> 101 papers received every grade from 1 to 9 on the scale; 94% from seven to nine different grades; and no essay received less than five different grades from the 53 expert readers.
>
> (Diederich 1974)

Additionally, when the rationales were examined for a factor analysis, it became apparent that the raters were looking at different factors or naming the same ones differently, weighting them differently in arriving at scores, and even disagreeing as to the nature and significance of errors.

> Definitions of proficiency in writing vary widely... with the [least] agreement at the upper rungs, where the stylistic preferences of teachers come into play. But even [then], there are disagreements about the importance of different errors and about the number of errors an educated reader will tolerate.

(Shaughnessy, 1977, 276)

Diederich's final scales consist of eight "clusters" under "General Merit" and "Mechanics": Topic, Ideas; Organization; Vocabulary; Style, Flavor; Language Use; Punctuation; Spelling; Format, Handwriting. "Ideas" and "Organization" were given double weight to satisfy the concerns of the teachers among the expert readers. The quality ranges from 1 to 5 points in each category, and this is to be determined holistically.

Primary trait scoring. "General Impression Marking" assigns a number to an essay, usually a composite derived from generally described categories. Papers receive no written comments or corrections. Lloyd-Jones comments, "The methods perfected by the ETS assume that excellence in one sample of one mode of writing predicts excellence in other modes--that is, good writing is good writing" (1977, 37). Dismissing the method as inadequate for failure to consider context, purpose, and intended audience, Lloyd-Jones developed Primary Trait Scoring under the auspices of the National Assessment for Educational Progress (1969-1970).

For the test, the writing task is structured narrowly and directions given to students emphasizing the most important considerations. Lloyd-Jones selects, for example, consistent point of view, use of dialogue, and control of tense as "primary traits" in a narrative. Levels of proficiency are limited to 0-2 or 0-3, unlike holistic scoring which often ranges from 5 to 10 points. Primary Trait Scoring allows the NAEP to compare groups by age in different writing tasks in both vertical and horizontal studies. With modification, it has potential for classroom application.

Revision and the Composing Process

Primary Trait Scoring offers no opportunity for revision; in

the test situation, the essay is a writing product and is summatively evaluated. Flower and Hayes ask, "How can evaluation change performance?...How can a teacher's response to a student's writing best help that writer improve?" (1979). Individual conferences are recommended in which questioning helps students to examine their own strategies and to find new ones (Murray, 1968; 1984). The importance of revision in improving writing is acknowledged by many (Emig, 1971, 1977; Gere, 1985; Nold, 1981) but convincing students of their need (and ability) to revise may be difficult (Odell & Cohick, "You mean, write it over in ink?" 1975).

Through between-draft evaluation, (Beach, 1979) many teachers attempt to intervene during the writing process, rather than waiting to react to the completed composition. A graded paper represents closure to students, and the editorial advice, corrections, and interlinear markings perceived as coming too late for the current assignment and premature for the next. (See especially Searle & Dillon, "The Message of Marking," 1980; Sommers on revision, 1980; and Camp, 1983, on involving students in evaluation). Peer group studies have typically found that when students are trained to give(and receive) criticism guided by a teacher-made revision checklist, they are given as much timely and usable feedback by their peers as by their teachers (Beaven, 1981; Benson, 1979; Lagana, 1972). Danis (1980) attributes peer review success to structured review sheets, in-class editing and feedback skills training, random assignment to peer groups, teacher involvement as resource and facilitator, and constant writing practice.

PURPOSE OF THE STUDY

The purpose of this study was to test the effects on student writing and attitudes toward grading, and on out-of-class teacher response time to papers, of a set of dual-purpose revision checklist/grading scales. The manner of presentation of the scales was the focus of the investigation: to determine whether direct teacher instruction or collaborative learning in peer critique groups was the more effective in communicating writing objectives and goals.

HYPOTHESES

It was expected that all students would demonstrate significant grade improvement as measured by the differences between pre- and posttest mean scores. Although teachers would likely report less out-of-class response time for peer-group papers, those students were expected to achieve as much or greater grade improvement as direct-instruction students. Additionally, student attitudes toward grading were expected to be comparable for the two groups.

LIMITATIONS

The study was conducted at the University of Maryland, College Park, in the Fall 1985 fifteen-week semester. Three Teaching Assistants with six sections of Freshman Composition volunteered for the study. The intact classes were arbitrarily assigned to treatment groups by their TA's, except for one ESOL group that was purposefully designated as a direct-instruction group. The fact that one of the TA's had been assigned one regular and one ESOL section of ENGL 101 came to light only after the teacher-training sessions had been completed.

METHODOLOGY

A set of analytic composition scales was developed through

modification of The Composition Evaluation Scales (Diederich, ETS, 1961) and Primary Trait Scoring (Lloyd-Jones, NAEP, 1969-1977). For each of the six major papers in Freshman Composition, an assignment sheet was coordinated with a grading scale. The scale was presented to all students to be used as a revision checklist after first drafts had been produced. Students in the direct-instruction group submitted their drafts to teachers and received written comments and in-class discussion as usual. Peer-group students had the opportunity to practice grading sample papers and classmates' drafts using the current scale. Teachers did not collect peer group drafts, but checked on peer grading sheets when the final copies were due.

Table 1

Sources of Data

Two independent variables were chosen as predictors of success: high school grade point average and score on a sample writing pretest (determined by independent raters using holistic guidelines from the ACE, 1985). The posttest was the final in-class writing assignment, scored under the same conditions and by the same raters as the pretest.

Teachers kept logs of grade breakdown for each assignment; these were not included in the study. A record of out-of-class time spent in responding to papers was also kept, and these data were processed to test one of the hypotheses. Student attitudes toward grading were elicited by means of end-of-the semester questionnaires and interviews.

Figure 1

Initial Procedures

Teachers were trained in holistic rating and in the constitution and preparation of peer critique groups. A first-week schedule for

Table 1

Differential Treatment

	Teacher Response Group	Peer Response Group
First Draft Due Date	Teacher collects, corrects, critiques drafts. No grade.	New scale discussed, applied to sample papers. Consensus reached on grades.
Next Class Session	Drafts returned, discussed. Checklist explained, taken home. Students begin revision in class, guided by written comments. Teacher available for individual advice, circulating around room.	Workshop: 2 peer editors critique & grade drafts. Teacher as facilitator. Drafts not collected. Checklists, marked by peers, guide revision.
Final Copy Due Date	Papers corrected & graded; checklists not resubmitted. Papers marked as usual; remediation prescribed by teacher.	Papers collected w/ peer sheets attached. Teacher marks, comments only on duplicate grade sheets. Papers remain unmarked. "Critical components" of minus categories circled for attention.

$$I \begin{cases} X_1 \text{---} P_1 \text{----} P_2 \text{----} (A_{1-5}) \text{----} A_6 \text{----} Q \text{----} T \\ X_2 \text{---} P_1 \text{----} P_2 \text{----} (A_{1-5}) \text{----} A_6 \text{----} Q \text{----} T \end{cases}$$

where I represents 6 intact sections of ENGL 101.

X_1 Direct-instruction group } arbitrarily designated by TA's

X_2 Peer-response group }

P_1 Pretest: in-class writing sample, 1st independent variable.

P_2 High school Grade Point Averages: 2nd predictor variable.

A_{1-5} Grade breakdowns for Assignments 1-5; not used in study.

A_6 Final in-class assignment used as posttest.

Q Student questionnaires with random personal interviews, at end of semester.

T Teachers' out-of-class response time as reported in Teacher Logs.

Figure 1. Research Design

Mon/Wed/Fri and Tue/Thu class sessions was produced for training peer group students. (See Week #1 Plan, Appendix A.)

Developing the Instrument

To apply the principles of behavioral objectives (Armstrong, R., Cornell, T., Kramer, R. & Roberson, E.W., 1970; Airasian, P. & Madaus, G., 1972; Kibler, R., Cegala, D., Barker, L. & Miles, D., 1974), the researcher intended that:

1. Criteria and standards were objectified for each assignment.
2. The high-performance level was described in each category.
3. The scoring system was explained so that students could use the scales independently from teacher instruction.
4. Drafts would be required and guidelines for revision given for each assignment throughout the semester.

Six common rhetorical patterns practiced in Freshman Composition were chosen for the study: description of a place; process explanation (to a child); division and classification; ad analysis through example; comparison/contrast (news event then and now); cause/effect (problem/solution). Six categories were selected for each paper to reflect logical organization, fluency of expression, and correctness "common to all good writing" (as measured by the Diederich scales). Additional categories specific to the rhetorical situation such as sensory detail in description or non-overlapping categories in classification represent "primary traits" (Lloyd-Jones). The set of six scales with coordinated assignment sheets is included as Appendices H & I in Boss (1986).

The categories were chosen to reflect cognitive, rhetorical, and linguistic competence. As mechanics of English were remediated throughout the semester (self-guided for peer-group students and suggested by teachers for the others), categories of "correctness"

were combined to allow for more complex criteria in later assignments. Criteria were derived by consulting both theorists and researchers (see especially Freedman, 1979, 1981; Halliday & Hasan, 1976; McQuade, 1979; Nystrand, 1982; Odell, L., Cooper, C. & Courts, C., 1978; Shuy, 1981a, 1981b). By contrast, see Olson's "grading slips" for an unsystematic selection of criteria and erratic scoring method (1982).

Scoring guide. A means of applying holistic scoring to analytical reading was found. Holistic scoring mandates that the features of text not be counted or deeply analyzed. This stipulation is offered by Cooper and Odell (1977). Yet some performance standards should be "kept in mind" as the grading proceeds, according to these researchers.

In each category of the researcher's scales, a plus (+) for superior, 2 points; or a check (✓) satisfactory, 1 point; or a minus (-) unacceptable, zero, is given. This is the extent of the discrimination, as in Lloyd-Jones' system. It involves even less deliberation than the five levels of the Diederich scales. Most of the analytic composition scales examined by the researcher (see Fagan et al., 1975) do not include scoring guides. Criteria-based scales are not sensitive to a quantitative continuum as are norm-referenced instruments. Instead, each category is assessed for presence or absence of that quality. Where directions for grading exist, they are complicated by the means of conversion to percentages and weighting of factors.

On the present scales, the maximum score is 12 points.(6 categories x 2 points). Points convert to letter grades as follows:

Table 2

Table 2

Conversion to Letter Grades

Points	Grade	Points	Grade	Points	Grade
12	A+	8	B	4	C-
11	A	7	B-	3	D+
10	A-	6	C+	2	D
9	B+	5	C	1	D-
				0	F*

*Students who received an F by not handing in the paper were not counted in the grade breakdown report; late papers were also excluded.

A sample assignment sheet and coordinated revision checklist/ grading scale are included as Appendix B.

RESULTS AND CONCLUSIONS

Technically, in this study, there was no "control" group. Both treatment groups received the identical checklists as guides to revision, and both were assured that teachers would use these and no other criteria in assigning grades. Differential treatment rested in the use of the scales and in the grade reporting method.

Hypotheses considered three areas: grade improvement as evidenced by holistic scoring of pre-test and posttest by independent raters; out-of-class response time reported by teachers; student attitudes toward grading procedures. The degree of teacher involvement in the revision of drafts (direct instruction) was measured through time spent at home or in the office in responding to individual papers. Total teacher response time included out-of-class conferencing, but did not include the in-class discussion of individual papers in the direct instruction group or the in-class role of facilitator for the peer critique group responding to drafts in workshop sessions.

Grade Improvement

Five sections of the six pretested close to the population mean (M = 6.0; SD = 3.89). Only the ESOL group was deviant, as expected, with a z score of -6.05 (deviation from the normal distribution of scores). This section exhibited the greatest pre-test/ posttest improvement, and only one section failed to achieve improvement far beyond probability levels.

Table 3

Table 3

Pretest/Posttest Mean Scores

Teacher Code	Group	N[a]	Pre-Test[b]	Post-Test[b]	S_X	S_Y	r	t-score
E X_1	T.R.[c]	18	5.6	7.5	2.0	1.2	.17	+3.8
G X_1	T.R.	12	3.7[d]	6.2	1.3	2.3	.74	+5.4
M X_1	T.R.	16	5.5	6.7	2.2	1.4	.38	+4.4
E X_2	P.R.[e]	15	5.7	6.8	2.3	1.8	.40	+1.9
G X_2	P.R.	19	5.9	6.5	1.3	2.3	.28	+1.2[f]
M X_2	P.R.	19	5.8	7.3	2.3	0.5	.29	+2.9

[a]Twelve students who had been pretested did not take post-test. These pretest scores were eliminated from the study.

[b]Six-point holistic ratings converted to 12-point researcher's scale.

[c]T.R. = Teacher Response Group (X_1).

[d]Deviant ESOL group pretested far below population mean (M = 6.0).

[e]P.R. = Peer Response Group (X_2).

[f]This section achieved no significant improvement in grades, based on critical t-ratio of $\pm$ 1.67, df 18, in a two-tailed test with $p < .05$.

Out-of-class Response Time

Teacher logs recorded response time to drafts (direct instruction group) and time spent in grading final papers for both groups. Peer critiques relieved the teachers of collecting drafts for that group, but peer grade sheets were submitted with the final papers so that teachers could monitor the amount and quality of feedback given. The method of grade reporting differed for the two groups as well. While guided by the researcher's scales in assigning grades, teachers corrected, commented, and gave editorial advice in writing on individual papers for the direct instruction group (as deemed necessary after thoroughly marking drafts).

Peer grading of drafts was reported only on the grading sheets; papers remained unmarked. A fresh grade sheet (identical to the revision checklist) was attached to the final copy, so that teachers also refrained from writing on the papers themselves.

Total average minutes for the semester per direct-instruction student was compared with that for each peer-group student. A two-tailed test of difference between two independent means was conducted (critical $z = \pm 1.96$; $p < .05$). An additional computation was made by dividing the total average minutes per student by the six assignments to find the average time spent per paper. Results were reported by teacher (section) and then by treatment group.

Table 4

Student Attitudes Toward Grading

Classroom observations were made periodically throughout the semester to check student reaction to the evaluation system. End-of-semester questionnaires, supported by individual random interviews,

Table 4

Average Total Out-of-class Response Time

Teacher Code	Group	Per Student for the Semester	Per Paper (Each of 6 Major Assignments)
E X_1	T.R.[a]	130 minutes	21 minutes
G X_1	T.R.	216 minutes[b]	36 minutes
M X_1	T.R.	155 minutes	26 minutes
E X_2	P.R.[c]	53 minutes	9 minutes
G X_2	P.R.	96 minutes	16 minutes
M X_2	P.R.	117 minutes	19 minutes
All Teacher Response		167 minutes	27 minutes
All Peer Response		89 minutes	14 minutes

[a]Teacher Response Group

[b]ESOL class

[c]Peer Response Group

z-score (the difference between two independent means) was found to be +2.38 (compare with z-critical $\pm$ 1.96, $p < .05$), a statistically significant difference.

gave a more systematic picture. Since there was no significant difference between the two groups, answers were pooled.

78% of all students usually or often felt that "requirements for 'A' papers were made clear for each assignment." 75% usually or often found that "the teacher did not show favoritism in grading." 70% "knew throughout the semester of my standing in class." 60% felt that "I was kept informed of my strengths as well as my problems." 53% believed that "my grades on final papers have been what I expected."

Yet in the peer-critique group, all three classes expressed anxiety about unmarked papers, the most common complaint that "written comments" were needed despite the specificity of the grading sheets. Teachers also were concerned that students were handicapped by not receiving thoroughly marked papers, with marginal notations and long editorial comments at the end. It became apparent through classroom observations and teacher/researcher conferences that "comments" actually meant "corrections," and that the students were objecting to accepting that responsibility for their own learning.

Although the revision checklists provided students with a rubric and a vocabulary for self- and peer-evaluation, teachers and students alike were reluctant to relinquish their traditional roles of information-giver and receiver. Teachers said that they felt "guilty," since grading went so quickly when guided by the predefined and described criteria on the checklist/grading sheets, and that they were not giving "equal time" to the peer-group students. All teachers were comfortable with the grading sheets as summaries of problems for remediation and as guides to lesson-planning.

FOR FURTHER STUDY

Since the present study did not attempt to validate the scales,

a new study might compare grades on individual papers awarded by the researcher's scales with those determined by Diederich's scales or pure holistic scoring (ACE Guidelines, 1985). The instrument itself was not assessed in the present study, which intended to integrate revision and grading into one system, to be practiced throughout the semester. Teacher Logs reported grade breakdowns for each assignment using the researcher's 12-point scale; however, this data was not submitted to statistical analysis.

Another variation can test the effect of advance knowledge of criteria (presentation of revision checklist/grading scale in advance of final copy submission) against the traditional procedure of grading final papers without presentation of specific criteria.

The scales might be used in conjunction with personal conferences in the experimental group while giving them as handouts without referring to them during control group conferences.

Grading might be standardized school- or district-wide on the high school level without affecting teaching or classroom practices.

The worthiness of such a follow-up project is summarized:

> ...to provide a student writer with a sense of audience, he must receive audience reactions <u>while</u> <u>engaged</u> in the process of writing, not at the end when the paper has been handed in, days have gone by, and the piece is handed back, minutely evaluated by the teacher.
>
> (Healy, in Camp, <u>Bay Area Writing Project</u>, 1983, 166, author's emphasis.)

REFERENCES

Airasian, P. & Madaus, G. Criterion-referenced testing in the classroom. Natl. Council on Measurement in Education, 3(4) 1-8, 1972.

American Council on Education. GED Scoring Guide. Available from ACE, 1 Dupont Circle, Wash., D.C. 20009, 1985.

Armstrong, R., Cornell, T., Kramer, R. & Roberson, E.W. The development and evaluation of behavioral objectives. Worthington, OH: Chas. A. Jones Publ., 1970.

Bata, E. A study of the relative effectiveness of marking techniques on junior college freshman English compositions. Unpublished PhD dissertation, Univ. of MD, College Park, 1972.

Beach, R. Effects of between-draft teacher evaluation vs. student self-evaluation on high school students' revision of rough drafts. Research in the Teaching of English, 13 (1979) 111-119.

Beaven, M.H. Individualized goal-setting, self-evaluation, and peer evaluation. In Cooper & Odell, Evaluating writing. Urbana, IL: Natl. Council of Teachers of English, 1977.

Benson, N. The effects of peer feedback during the writing process on writing performance, revision, and attitude toward writing. Unpublished PhD dissertation, Univ. of CO at Boulder, 1979. Available from Univ. Microfilms, Ann Arbor, MI. #AAD 79-23-212.

Bloom, B., Hastings, T. & Madaus, G. Handbook on formative and summative evaluations of student learning. NY: McGraw-Hill, 1971.

Boss, R. Peer group critiques in formative evaluation of college composition. Unpublished PhD dissertation, Univ. of MD, College Park, 1986.

California Assn. of Teachers of English. A scale for evaluation of high school student essays. Urbana, IL: Natl. Council of Teachers of English, 1960.

Camp, G. (Ed.) Teaching writing: Essays from the Bay Area Writing Project. Montclair, NJ: Boynton/Cook Publ., 1983.

Charney, D. The validity of using holistic scoring to evaluate writing: A critical overview. Research in the Teaching of English, 18:1 (Feb. 1984) 65-81.

Composition Evaluation Scales. Developed by Diederich, P., French, J. & Carlton, S. Factors in judgments of writing ability. Princeton, NJ: Educational Testing Service, 1961. Scales are available from ERIC #ED 091-750.

Cooper, C. & Odell, L. (Eds.) Research in composing: Points of departure. Urbana, IL: Natl. Council of Teachers of English, 1978.

Danis, M.F. Peer-response groups in a college writing workshop: Students' suggestions for revising compositions. Unpublished PhD dissertation, Michigan State Univ., 1980. Available from Univ. Microfilms, Ann Arbor, MI. #AAD 81-12-066.

Diederich, P. Measuring growth in English. Urbana, IL: Natl. Council of Teachers of English, 1974.

Fagan, W., Cooper, C. & Jensen, J. Measures for research and evaluation in English language arts. Urbana, IL: Natl. Council of Teachers of English, 1975.

Flower, L. & Hayes, J. The cognition of discovery: Defining a rhetorical problem. College Composition and Communication, 30 (1980) 26-32.

Emig, J. Writing as a mode of learning. College Composition & Communication, 18 (1977) 122-128.

Freedman, S. Influences on the evaluation of expository essays: Beyond the text. Research in the Teaching of English, 15 (1981) 245-255.

Freedman, S. & Calfee, R. Holistic assessment of writing: Experimental design and cognitive theory. In Mosenthal, Tamor & Walmsley

(Eds.) Research on writing: Principles and methods. NY: Longman Publ., 1983.

Gere, A. Writing and learning. NY: Macmillan, 1985.

Gere, A. Written composition: Toward a theory of evaluation. College English, 42 (1980) 44-58.

Godshalk, F., Swinford, F. & Coffman, W. The measurement of writing ability. NY: College Entrance Examination Board, 1966.

Halliday, M. & Hasan, R. Cohesion in English. London: Longman Group, 1976.

Healy, M.H. On small groups. In Camp, G. (Ed.) Teaching writing: Essays from the Bay Area Writing Project. Montclair, NJ: Boynton/Cook Publ., 1983.

Kibler, R., Cegala, D., Barker, L. & Miles, D. Objectives for instruction and evaluation. Boston: Allyn & Bacon, 1974.

Lagana, J.R. The development, implementation, and evaluation of a model for teaching composition which utilizes individualized learning and peer grouping. Unpublished PhD dissertation, Univ. of Pittsburgh, 1972. Available from Univ. Microfilms, Ann Arbor, MI: #AAD 73-04-127.

Lloyd-Jones, R. Primary trait scoring. In Cooper & Odell, Evaluating writing. Urbana, IL: Natl. Council of Teachers of English, 1977.

Lloyd-Jones, R. Rhetorical choices in writing. In Frederiksen & Dominic (Eds.) Writing, 2. Hillsdale, NJ: Erlbaum Assoc., 1981.

McQuade, D. & Atwan, R. Thinking in writing. NY: Knopf, 1980.

Murray, D. Write to learn. NY: Holt, Rinehart & Winston, 1984.

Murray, D. A writer teaches writing. Boston: Houghton-Mifflin, 1968.

National Assessment for Educational Progress. Explanatory and persuasive letter-writing, 1977; Expressive writing, 1976; Writing mechanics, 1969-74. Denver, NAEP.

Nold, E. Revising. In Frederiksen & Dominic (Eds.) Writing, 2. Hillsdale, NJ: Erlbaum Assoc., 1981.

Nystrand, M. What writers know: The language process and structure of written discourse. NY: Academic Press, 1982.

Odell, L. & Cohick, J. You mean, write it over in ink? English Journal, 64 (Dec. 1975) 48-53.

Odell, L., Cooper, C. & Courts, C. Discourse theory: Implications for research in composing. In Cooper & Odell (Eds.) Research on composing: Points of departure. Urbana, IL: Natl. Council of Teachers of English, 1978.

Olson, M. The writing process: Composition and applied grammar. Boston: Allyn & Bacon, 1982.

Searle, D. & Dillon, D. The message of marking. Research in the Teaching of English, 14 (1980) 233-242.

Shaughnessy, M. Errors and expectations. NY: Oxford Univ. Press, 1977.

Shuy, R. A holistic view of language. Research in the Teaching of English, 15 (1981a) 101-112.

Shuy, R. Toward a developmental theory of writing. In Frederiksen & Dominic (Eds.) Writing, 2. Hillsdale, NJ: Erlbaum Assoc., 1981b.

Sommers, N. Revision strategies of student writers and experienced adult writers. College Composition and Communication, 31 (Dec. 1980) 378-388.

Stanton, B.E. A comparison of theme grades written by students possessing varying amounts of cumulative written guidance: Checklist, instruction, and questions/feedback. Unpublished EdD dissertation, Brigham Young Univ., 1974.

Stevens, A.E. The effects of positive and negative evaluation on the written composition of low performing high school students. Unpublished EdD dissertation, Boston Univ., 1973.

Appendix A

Week #1 Plan for Peer Group

(MWF 50-minute sessions; TuTh 75-minute sessions; Total 150 minutes)

Minutes	Activity	Purpose
As students enter 5	Pick up, begin filling out Student Information Sheets.	To provide teacher with data for forming heterogeneous peer groups.
15	Hand out & refer to syllabus, Course Policy Statements, grammar test review, other forms. Students skim as Info Sheets collected.	To handle administrative tasks required by Engl. Dept. To assign students to peer groups.
10	Students individually make 10-item list of most important components of good writing.	To involve students immediately in determining criteria for evaluation.
10	Students assemble in peer groups to discuss criteria & reach consensus to refine list.	To convince students of their own prior knowledge of writing criteria.
10 End of 1st MWF class	Each group elects spokesperson to present its list to class & a secretary to write on board.	To begin collaborative effort in a writing task. To make oral defense of group decision.
15	Teacher summarizes student criteria; presents	To relate researcher's criteria to students' own. To

APPENDIX A.2		Formative Evaluation
	combination revision checklist/grade scale concept.	introduce peer critique workshops & formative evaluation methods.
5	Teacher demonstrates grading points & holistic rating, conversion to letter grades.	To prepare for homework assignment: student grading of sample paper.
5 End of 1st Tu/Th class	Peer group schedules response to papers for each assignment.	To ensure 2 peer editors per paper & rotation of reviewers in 5-member group.
20	Writing Sample administered, to be graded holistically by independent raters.	For 1st independent variable (comparability of groups before instruction) pretest.
5 End of 2nd MWF class	Teacher demonstrates use of Writing Sample; gives out Student Permission letters.	To allow researcher access to certain student data.
20	Presentation of results of grading sample paper. Discussion of criteria & use of info. for revision.	To have students reach consensus on quality of writing (analytic categories scored holistically).
15	Grammar review for Engl. Proficiency Examination.	Unrelated to study; admin. task required by Engl. Dept.

APPENDIX A.3		Formative Evaluation
15	Student reaction paper: Relation of Grammar Study to Composing.	To make students aware of "content vs. correctness" controversy in grading.

****

End of 3rd MWF class; 2nd Tu/Th class; Total 150 minutes.

Appendix B.1

Sample Assignment Sheet

Assignment 4 (of 6): Ad Analysis through Example

Skills: To identify a brand-name product image, lifestyle, in an ad.

To find appeals in each ad, separating facts from claims.

To characterize audience for each ad; magazine readership.

Intended Reader: A friend who is very brand-conscious.

Prewriting: Two in-class exercises provide the "data base" for this assignment. In class discussion, you will see how ads are designed to affect us on various levels of appeal, in terms of ethos, logos & pathos. Playing "The Ad Game" will show you that products are named to trigger emotions. The worksheets that you make will show how different brands of the same product can be given an "image" to suit the intended reader's "lifestyle" or fantasies.

Procedure: Find 3 ads for 3 different brands of the same product. We will practice with cigarette ads, because there is very little logic involved in smoking. The ads must appear in 3 very different magazines, preferably those you have no interest in, so that you can get the distance necessary for analysis. Make a worksheet for each ad. First,write a typical reader profile by examining the cover, article titles, subjects of features, and mix of ads. Then find all the information you can about the product advertised, including the model, props, scenery, logo, slogan, colors, amount of copy. List 3 facts and 3 claims (remember our "Fact vs. Opinion" exercise).

Write thesis statement about how advertising works. Let the reader know "where you're coming from." The stronger your feelings, the better will be your argument. Make a topic outline to show how you will organize examples from the magazine and from the ad itself to

prove your points about advertising. Show why this ad is effective; how it "works" on its target audience. Do not compare all 3 ads; take all of your examples from the best ad of the 3 you chose. Write your first draft for the workshop. Be sure to include your worksheet and the ad itself, but not the whole magazine.

APPENDIX B.3 Formative Evaluation

Sample Revision Checklist/Grading Sheet

Name_____ENGL 101 Section_____Date_____

Check one: First draft____Final copy____If draft for revision workshop, list names of peer reviewers:_____

The high level of each category is described (+ for 2 points). A check for satisfactory (✓= 1 point) or minus for unacceptable (- = 0) are the other choices. If zero is given, please circle the critical components that apply. Write comments on back of sheet; DO NOT MARK OR CORRECT PAPERS. Use scoring guide for conversion of points to letter grades.

#4: AD ANALYSIS THROUGH EXAMPLE Total Points____ Letter grade____

1. Observation: Descriptive, specific details of ad elements such as color, copy, layout, model. Quality___ Points____
2. Objectivity: Facts & inferences about magazine reader characteristics, like age, socioeconomic level, status needs (stereotypes). Quality___ Points____
3. Analysis: Exposure of "hidden" psychological appeals (like trigger words or attractive model). Pay attention to logical fallacies uncovered here. Quality___ Points____
4. Logical Progression: Organized so that statements are supported by evidence from ad & magazine; use of paragraphs & transitions to reflect the drawing of conclusions from details. Quality___ Points____
5. Word Choice: Precision in vocabulary, especially in describing the ad. Correct diction, usage; terms defined. Quality___ Points____
6. Mechanics/Sentence Structure: Correctness in spelling, agreement, tense, punctuation, essay form, pronoun use. Quality___ Points____

ERIC CLEARINGHOUSE FOR JUNIOR COLLEGES FEB 19 1988